



(19)

(11) Publication number:

1

Generated Document.

PATENT ABSTRACTS OF JAPAN(21) Application number: **10007321**(51) Intl. Cl.: **G06F 3/06 G06F 3/06**(22) Application date: **19.01.98**

(30) Priority:	
(43) Date of application publication: 30.07.99	(71) Applicant: FUJITSU LTD
(84) Designated contracting states:	(72) Inventor: TAKEDA SUJIN
	(74) Representative:

**(54) INPUT/OUTPUT
CONTROLLER AND ARRAY
DISK DEVICE**

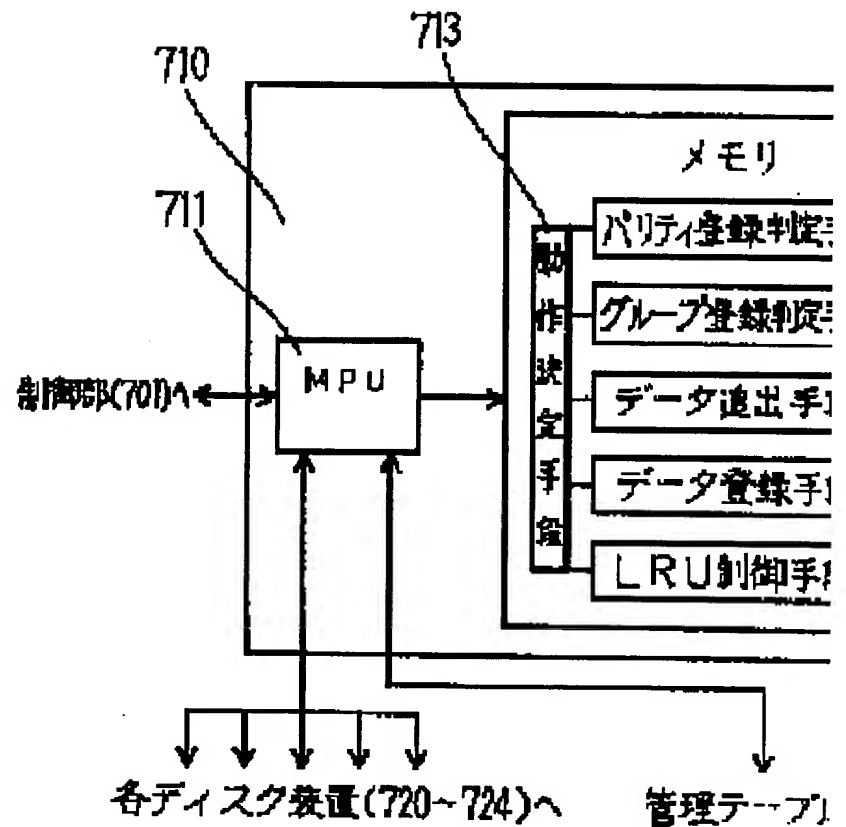
(57) Abstract:

PROBLEM TO BE SOLVED: To eliminate deceleration of access speed and to enable high-speed access by providing a cache managing part with a group register means for discriminating whether the other data consisting of a parity group, to which the data of access object belong, are registered on a cache memory or not.

SOLUTION: A group registration discriminating means 717 of a cache managing part 710 judges whether the other data consisting of the parity group, to which the object data belong, are already registered in the cache memory or not. The group registration discriminating means 717 refers to an entry table in a managing table. The group registration discriminating means 717 retrieves the logical block number of data as

the access object out of 'object blocks' in the entry table. When the object logical block is detected, the parity data of the parity group, to which the object data belong, are already registered in the cache memory but when such a block is not detected, these data are not registered.

COPYRIGHT: (C)1999,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-203056

(43) 公開日 平成11年(1999) 7月30日

(51) Int.Cl.⁵
G 0 6 F 3/06

識別記号
5 4 0
3 0 2

F I
G 0 6 F 3/06

5 4 0
3 0 2 A
3 0 2 Z

審査請求 未請求 請求項の数 8 O L (全 22 頁)

(21) 出願番号 特願平10-7321

(22) 出願日 平成10年(1998) 1月19日

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72) 発明者 武田 帥仁

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

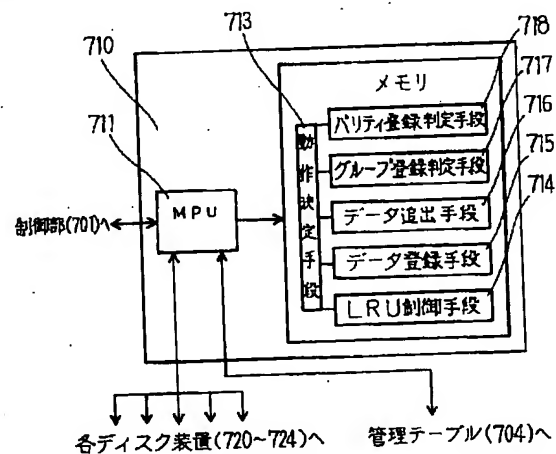
(74) 代理人 弁理士 井桁 貞一

(54) 【発明の名称】 入出力制御装置及びアレイディスク装置

(57) 【要約】

【課題】 ディスク待ち時間によるアクセス速度の低下をなくし、信頼性が高く、かつ高速アクセスを可能とするRAID4及びRAID5型のディスク装置を実現すること。

【解決手段】 キャッシュ管理部(710)はアクセス対象となったデータが属するパリティグループを構成する他のデータがキャッシュメモリ上に登録されているかを判定するグループ登録判定手段(717)を含むこと、を特徴とする入出力制御装置を提供する。



【特許請求の範囲】

【請求項1】複数の入出力装置に接続され、当該入出力装置に格納されているデータを記憶するキャッシュメモリと、

該キャッシュメモリに記憶されているデータが格納されるべき前記入出力装置上の位置を示すキャッシュメモリ管理テーブルと、

前記キャッシュメモリ管理テーブルを用いて前記入出力装置と当該キャッシュメモリとの間でのデータ転送を制御するキャッシュ管理部と、

上位装置からの要求に応じて、前記キャッシュ管理部の判断に基づき上位装置とキャッシュメモリとの間のデータ転送を制御する制御部と、を備えた入出力制御装置であって、

前記キャッシュ管理部は、アクセス対象となったデータが属するパリティグループを構成する他のデータがキャッシュメモリ上に登録されているか否かを判定するグループ登録判定手段を含むこと、を特徴とする入出力制御装置。

【請求項2】請求項1記載の入出力制御装置であって、前記キャッシュ管理部は、アクセス対象となったデータを基礎として作成された冗長データがキャッシュメモリ上に登録されているか否かを判定するパリティ登録判定手段を含むこと、を特徴とする入出力制御装置。

【請求項3】請求項1記載の入出力制御装置であって、前記キャッシュ管理部は、アクセス頻度に応じてキャッシュメモリ上に登録されているデータの優先順位を決定するという制御を行うLRU制御手段を含み、該LRU制御手段は、キャッシュメモリ上に登録されたパリティデータを当該LRU制御の対象としないこと、を特徴とする入出力制御装置。

【請求項4】請求項1記載の入出力制御装置であって、前記キャッシュメモリ管理テーブルは、冗長データと当該冗長データを作成する基礎となった全てのデータとの対応を示す情報を備えること、を特徴とする入出力制御装置。

【請求項5】請求項1記載の入出力制御装置であって、前記キャッシュ管理部は、キャッシュメモリ上に登録されているある冗長データを作成する基礎となったデータのいずれかがキャッシュメモリ上に登録されている限り当該冗長データをキャッシュメモリ上に登録し続ける機能を備えたデータ追出し手段を含むこと、を特徴とする入出力制御装置。

【請求項6】請求項1記載の入出力制御装置であって、前記キャッシュ管理部は、冗長データを作成する基礎となったデータのいずれかをキャッシュメモリ上に登録する場合には当該冗長データも登録する機能を備えたデータ登録手段を含むこと、を特徴とする入出力制御装置。

【請求項7】請求項1記載の入出力制御装置であって、

前記キャッシュ管理部は、前記冗長データと該冗長データを作成する基礎となったデータとの全てを一体としてキャッシュメモリへの登録を管理する機能を備えたデータ登録手段を含むこと、を特徴とする入出力制御装置。

【請求項8】上位装置との間で転送されるデータが分散して格納されるとともに、該分散されたデータに基づいて作成される冗長データが格納される複数の入出力装置と、

該複数の入出力装置に接続され、当該入出力装置に格納されているデータを記憶するキャッシュメモリと、

該キャッシュメモリに記憶されているデータが格納されるべき前記入出力装置上の位置を示すキャッシュメモリ管理テーブルと、

前記キャッシュメモリ管理テーブルを用いて前記入出力装置と当該キャッシュメモリとの間でのデータ転送を制御するキャッシュ管理部と、

上位装置からの要求に応じて、前記キャッシュ管理部の判断に基づき上位装置とキャッシュメモリとの間のデータ転送を制御する制御部と、を備えたアレイドиск装置であって、

前記キャッシュ管理部は、キャッシュメモリ上に登録されているある冗長データを作成する基礎となったデータのいずれかがキャッシュメモリ上に登録されている限り当該冗長データをキャッシュメモリ上に登録し続ける機能を備えたデータ追出し手段を含むこと、を特徴とするアレイドиск装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、複数のディスク装置を並列にアクセスしてデータ入出力処理を行う入出力装置及びアレイドиск装置に関するものである。特にディスク装置に格納したデータを更新する場合の性能を向上させた入出力制御装置及びアレイドиск装置に関する。

【0002】

【従来の技術】従来、カリフォルニア大学バークレイ校のデビッド・A・バターソン(David A. Patterson)らは、大量のデータを多くのディスクに転送することで、高速かつ高いデータの信頼性を実現するディスクアレイド装置について、レベル1から5までに分類付けを行って評価した論文を発表している(ACMS IGMOD Conference, Chicago, Illinois, June 1-3, 1988P109-P166)。

【0003】このデビッド・A・バターソンらが提案したディスクアレイド装置を分類する際に用いられる、レベル1から5という概念は、RAID(Redundant Array of Inexpensive Disks)1から5と略称される。本発明は、以上のRAIDレベルによる分類においては、特にRAID4及びRAID5に関するものである。

【0004】図16は、従来のキャッシュ管理部(710)の構造を示したものである。以降、全図を通して同一部分には同一番号を付す。

キャッシュ管理部(710)は、MPU(711)及びメモリ(712)から構成されている。MPUは、メモリ、管理テーブル(704)、各ディスク装置(720~724)及び制御部(701)と通信可能に接続されている。メモリにはMPUの動作を制御するためのマイクロプログラムが格納されている。MPUは、このマイクロプログラムによって記述された制御論理に従って管理テーブルや制御部との通信を行い、キャッシュメモリ上に登録されているデータの管理を行う。メモリに格納されているマイクロプログラムは、動作決定手段(713)、LRU制御手段(714)、データ登録手段(715)及びデータ追出手段(716)を含むものである。

【0005】動作決定手段は制御部からの通信を検知する。また、キャッシュ管理部が行うべき処理を決定し各処理を実行する個々の手段を起動する。動作決定手段は、ディスク装置及び上位装置から転送されたデータをキャッシュメモリ上に登録する必要があるときはデータ登録手段を起動する。データ登録手段は、キャッシュメモリに格納されたデータを有効にするためにキャッシュメモリ(703)上の管理テーブル(704)を編集する。そしてデータ登録手段は、当該データを登録済みの状態にする。次に、動作決定手段はLRU制御手段を起動する。LRU制御手段は登録されたデータのキャッシュメモリ内における優先順位を決定し、管理テーブルに反映する。

【0006】動作決定手段は、既に使用されているキャッシュメモリ等を解放する必要が生じた場合は、まずLRU制御手段を起動する。LRU制御手段は管理テーブルを参照する。これにより、LRU制御手段はキャッシュメモリ上に登録されているデータの優先順位を認識する。その結果、LRU制御手段はキャッシュメモリ上から登録を抹消すべきデータを決定する。動作決定手段は、LRU制御手段が決定したデータを抹消するために、データ追出し手段を起動する。データ追出し手段は、管理テーブルを編集する。そしてデータ追出し手段は、キャッシュメモリに格納されているデータを未登録状態にする。

【0007】また、動作決定手段は、アクセス対象となるデータがキャッシュメモリ上に登録されているか否かの問い合わせを制御部から受けると、LRU制御手段を起動する。LRU制御手段は、管理テーブルを参照して対象データがキャッシュメモリ上に登録されているか否かを判断する。そして、LRU制御手段は処理を動作決定手段に戻す。動作決定手段は、LRU制御手段が判断した対象データの状態を制御部に通知する。

【0008】

【発明が解決しようとする課題】以上のように、RAID4あるいはRAID5技術は、ディスク装置の信頼性を飛躍的に高めることを可能とする。しかし、そのためには、パリティデータを常に正しい値に保っておく必要がある。つまり、あるデータが書き換えられた場合には、当該データが属するパリティグループのパリティを再計算し、当該データの格納と同時にパリティデータの更新も行わなければならない。このため、データを書き込む際には、当該データを含む記憶ブロックへアクセスを行うだけでなく、パリティデータを含む記憶ブロックへのアクセスをも必要とする。

【0009】したがって、本発明における形態のディスク制御装置においては、書込みに際して、必ず複数のディスク装置へのアクセスが必要となる。つまり、ディスク装置の位置付け待ち及び回転待ちが複数のディスク装置に対して生ずる。このため、従来型のディスク装置に比べて、RAID4あるいはRAID5はディスク装置の待ち時間が増加するのである。つまり、RAID4及びRAID5では、書込み処理においては、大幅な速度低下を起こすこととなる。

【0010】本発明は、このようなディスク待ち時間によるアクセス速度の低下をなくし、信頼性が高く、かつ高速アクセスを可能とするRAID4及びRAID5型のディスク装置を実現することを目的とする。

【0011】

【課題を解決するための手段】以上のような問題を解決するために、請求項1に記載の発明は、図1に示すように、複数の入出力装置に接続され当該入出力装置に格納されているデータを記憶するキャッシュメモリ(703)と、該キャッシュメモリに記憶されているデータが格納されるべき前記入出力装置上の位置を示すキャッシュメモリ管理テーブル(704)と、前記キャッシュメモリ管理テーブルを用いて前記入出力装置と当該キャッシュメモリとの間でのデータ転送を制御するキャッシュ管理部(710)と、上位装置(730)からの要求に応じて前記キャッシュ管理部の判断に基づき上位装置とキャッシュメモリとの間のデータ転送を制御する制御部(701)と、を備えた入出力制御装置(700)であって、前記キャッシュ管理部(710)はアクセス対象となったデータが属するパリティグループを構成する他のデータがキャッシュメモリ上に登録されているか否かを判定するグループ登録判定手段(717)を含むこと、を特徴とする入出力制御装置を提供する。

【0012】本請求項に記載された発明にあっては、キャッシュメモリに登録されているデータを追い出す場合には、当該データが属するパリティグループを構成する他のデータがキャッシュメモリ上に登録されている場合にはパリティデータをキャッシュメモリ上に維持することができる。また、当該他のデータがキャッシュメモリ上に存在しない場合には、パリティデータも追い出し対象

データと同時に追い出すことができる。

【0013】請求項2に記載された発明は、前記キャッシュ管理部が更に図1に示すアクセス対象となったデータを基礎として作成された冗長データがキャッシュメモリ上に登録されているか否かを判定するパリティ登録判定手段(718)を含むこと、を特徴とする入出力制御装置を提供する。本請求項に記載された発明にあっては、既にキャッシュメモリに登録されているパリティデータについては、再度のデータ転送を抑止することができる。

【0014】請求項3に記載された発明は、前記キャッシュ管理部が更に図1に示すようにアクセス頻度に応じてキャッシュメモリ上に登録されているデータの優先順位を決定するという制御を行うLRU制御手段(714)を含み、該LRU制御手段はキャッシュメモリ上に登録されたパリティデータを当該LRU制御の対象としないこと、を特徴とする入出力制御装置を提供する。

【0015】本請求項に記載された発明にあっては、パリティデータを一般のユーザデータと区別して取り扱うことができる。請求項4に記載された発明は、前記キャッシュメモリ管理テーブルが冗長データと当該冗長データを作成する基礎となった全てのデータとの対応を示す情報を備えること、を特徴とする入出力制御装置を提供する。

【0016】本請求項に記載された発明にあっては、請求項1に記載された発明と同様に、キャッシュメモリに登録されているデータを追い出す場合に、当該データが属するパリティグループを構成する他のデータがキャッシュメモリ上に登録されている場合にはパリティデータをキャッシュメモリ上に維持することができる。請求項5に記載された発明は、前記キャッシュ管理部が更に図1に示すキャッシュメモリ上に登録されているある冗長データを作成する基礎となったデータのいずれかがキャッシュメモリ上に登録されている限り当該冗長データをキャッシュメモリ上に登録し続ける機能を備えたデータ追出し手段(716)を含むこと、を特徴とする入出力制御装置を提供する。

【0017】本請求項に記載された発明も請求項1に記載された発明と同様に、キャッシュメモリに登録されているデータを追い出す場合に、当該データが属するパリティグループを構成する他のデータがキャッシュメモリ上に登録されている場合にはパリティデータをキャッシュメモリ上に維持することができる。請求項6に記載された発明は、前記キャッシュ管理部が、更に図1に示す冗長データを作成する基礎となったデータのいずれかをキャッシュメモリ上に登録する場合には当該冗長データも登録する機能を備えたデータ登録手段(715)を含むこと、を特徴とする入出力制御装置を提供する。

【0018】本請求項に記載された発明にあっては、予めパリティデータをキャッシュメモリ上に登録しておく

ことができる。このため、上位装置から書込み処理を要求された際にディスク装置へのアクセスを行わないで済む。請求項7に記載された発明は、前記キャッシュ管理部が、更に図1に示す前記冗長データと該冗長データを作成する基礎となったデータとの全てを一体としてキャッシュメモリへの登録を管理する機能を備えたデータ登録手段(715)を含むこと、を特徴とする入出力制御装置を提供する。本請求項に記載された発明にあっては、あるパリティグループに含まれる全てのデータを一群のデータとして管理できる。

【0019】請求項8に記載された発明は、上位装置との間で転送されるデータが分散して格納されるとともに該分散されたデータに基づいて作成される冗長データが格納される複数の入出力装置と、該複数の入出力装置に接続され当該入出力装置に格納されているデータを記憶するキャッシュメモリと、該キャッシュメモリに記憶されているデータが格納されるべき前記入出力装置上の位置を示すキャッシュメモリ管理テーブルと、前記キャッシュメモリ管理テーブルを用いて前記入出力装置と当該キャッシュメモリとの間でのデータ転送を制御するキャッシュ管理部と、上位装置からの要求に応じて前記キャッシュ管理部の判断に基づきディスク装置とキャッシュメモリとの間のデータ転送を制御する制御部と、を備えたアレイドスク装置であって、前記キャッシュ管理部はキャッシュメモリ上に登録されているある冗長データを作成する基礎となったデータのいずれかがキャッシュメモリ上に登録されている限り当該冗長データをキャッシュメモリ上に登録し続ける機能を備えたデータ追出し手段を含むこと、を特徴とするアレイドスク装置を提供する。

【0020】本請求項に記載された発明にあっては、請求項1に記載された発明と同様に、キャッシュメモリに登録されているデータを追い出す場合に、当該データが属するパリティグループを構成する他のデータがキャッシュメモリ上に登録されている場合にはパリティデータをキャッシュメモリ上に維持することができる。

【0021】

【実施の形態】A. 入出力装置の全体構成の説明

図7は、本発明が適用されるディスク制御装置の説明図である。ディスク制御装置(700)には、複数のディスク装置(720~724)が接続されている。各ディスク装置には、上位装置(730)との間で転送されるデータが一定長ずつ分散されて格納されている。

【0022】また、ディスク制御装置(700)は、キャッシュメモリ(703)と、キャッシュ管理部(702)と、制御部(701)と、から構成されている。キャッシュメモリには、ディスク装置(720~724)に格納されているデータの一部及び、キャッシュメモリ(703)上に記憶されているデータと当該データが格納されるべき前記ディスク装置内の位置関係を示す管理

テーブル(704)から成る。

【0023】キャッシュ管理部(702)は、前記管理テーブルの参照や更新を行う。また、キャッシュ管理部はキャッシュメモリ(703)上に記憶されるデータの管理及び制御も行う。制御部(701)は、上位装置(730)からの要求に応じて、キャッシュメモリ(703)と上位装置との間でのデータ転送を行う。

【0024】図7では、一例として上位装置との間で転送されるデータが五つのディスク装置に分散されて格納されている場合を示している。ただし、五つのディスク装置(720~724)のそれぞれについて、同一のアドレスで示されるブロックに格納されているデータに着目した場合、実際に上位装置との間で転送されるデータが格納されているディスク装置は四つである。そして、残りの一つのディスク装置には前記四つのディスク装置に格納されたデータから作成されるパリティデータが格納されている。

【0025】これを図8において模式的に示す。図8は図7で示した五つのディスク装置のデータ格納状態を示している。それぞれのディスク装置(Disk0~Disk4)には五つの記憶ブロックが存在し、それぞれBlock0~Block4と呼ぶ。したがって、これら複数のディスク装置内のブロックは、ディスク装置番号とブロック番号を用いることで特定することができる。このディスク装置番号とブロック番号を用いる対象データの指定方法を物理アドレスによる指定と呼ぶ。

【0026】ところで、ディスク制御装置は、上位装置とのインタフェース上の互換性を保つ必要がある。したがって、当該ディスク制御装置は一つの大容量ディスク装置のみが接続されているが如く動作する必要がある。このため、上位装置からの、それぞれの記憶ブロックに対するアクセスは、上位装置が認識しているディスク装置、すなわち制御装置がエミュレーションしている仮想的なディスク装置に対する番号と当該仮想的なディスク装置内の仮想的なブロック番号で指定される。このような上位装置が用いる特定ブロックの指定方法を論理アドレスによる指定と呼ぶ。

【0027】論理アドレスは、通常0から1づつ昇順に存在する。前述したようにディスク制御装置は、上位装置から指定されるアドレスを論理アドレスと認識する。そして、実際にディスク制御装置にアクセスする際には、論理アドレスを物理的地址、すなわち物理ディスク装置番号と当該物理ディスク装置内の物理ブロック番号に変換して、アクセスを行う。

【0028】各物理ディスク装置Disk0からDisk4には、それぞれ記憶ブロックが存在する。ある記憶ブロックを一意に指定する場合は、ディスク番号とブロック番号が必要である。例えば、LB-6を指定する場合は、物理アドレスDisk2/Block2を指定する。しかし、ディスク制御装置は、上位装置にはDisk0~Disk4の複数のディスク装置を論理的に一台のディスク装置として認識させなければならない。このため、ディスク制御装置は、ディスク制御装置上で論理アドレスを物理アドレス、すなわち物理ディスク番号と当該物理ディスク装置におけるブロック番号との組み合わせに変換する必要がある。

【0029】ディスク制御装置は、この論理アドレスを装置内で物理アドレスに変換して目的のデータにアクセスしている。以下、この変換方法の一例を示す。ただし、説明を簡単にするため、当該ディスク制御装置は、一つの仮想的なディスク装置のみ存在しているが如く動作するものとする。従って、以下の説明では、仮想的なディスク番号については説明を省略する。

【0030】論理アドレスから物理アドレスへの変換は、接続されている物理ディスク装置の台数に基づいて行われる。つまり、対象となる物理ディスク装置の番号は、以下の値を用いて決定することができる。

値A = {論理アドレス / (物理ディスク装置台数 - 1)} の余り
値B = {論理アドレス / (物理ディスク装置台数 - 1)} の小数点以下を切り捨てた値
値C = (値B / 物理ディスク装置台数) の余り

とすると物理ディスク番号は、図9に示した表から求められる。

【0031】また、対象となる物理ブロック番号は、前述した式により求められる値Bである。具体的には、図8の様な構成であれば、接続されている物理ディスク装置の台数が5台であるから、論理アドレスから物理アドレスを求めるには以下の値を用いる。

値A = {論理アドレス / (5 - 1)} の余り
値B = {論理アドレス / (5 - 1)} の小数点以下を切り捨てた値
値C = (値B / 5) の余り

例えば、論理アドレスが11であれば、値A = 3、値B = 2、値C = 2、となるから図9の表よりディスク4が決定される。物理アドレスはディスク4のブロック番号2となる。

【0032】ところで、五つのディスク装置の同一ブロック番号で示される領域に着目した場合、以上の変換によれば、四つのディスク装置には上位装置との間で転送されるデータが格納されるが、残りの一つにはデータが割当てられないこととなる。このデータが割当てられないブロックには、他のディスク装置の同一ブロック番号に格納されているデータから作成されるパリティデータが格納される。このパリティデータが存在することにより本発明にかかるディスク装置は、その冗長性、ひいては信頼性を向上させている。

【0033】また、パリティが格納される物理ディスクは、以下の式から求められる。

値D = (物理ディスク装置台数 - 1 - 値C)

図8に示した構成であれば、パリティが格納されている物理ディスクは、

値 $D = (4 - \text{値}C)$

となる。例えば論理アドレス11を要素とするパリティが格納される物理ディスク番号はDisk2である。

【0034】ここで、あるパリティデータを作成するために用いられるデータ群と当該パリティとを含めた集合をパリティグループと呼ぶ。図8を用いて例示すれば、物理アドレスDisk0/Block0と、Disk1/Block0と、Disk2/Block0と、Disk3/Block0と、に含まれるデータから作成されたパリティデータP0が、物理アドレスDisk4/Block0に格納されている。そして、以上の五つのブロックが一つのパリティグループを形成していることとなる。

【0035】Disk0/Block1からDisk4/Block1についても、これと同様に上位装置との間で転送されるデータが分散格納される。ここで、パリティデータP1に関しては本実施例においてはディスク装置3に格納されるものとしている(RAID5)。これは便宜上のものであって、もちろんパリティデータを格納するディスクを固定して、常にDisk4に格納することとしても差し支えない(RAID4)。

B. キャッシュ管理部の説明

図2は、キャッシュ管理部(710)の動作を示したフローチャートである。キャッシュ管理部は、通常は制御部(701)からの通信待の状態にある(S201)。この状態は、動作決定手段にしたがってMPUが動作している状態である。動作決定手段に基づいて動作しているキャッシュ管理部は、制御部からの通信を検出すると(S202)、処理をLRU制御手段(714)に移す。LRU制御手段は、制御部から通知されたデータがキャッシュメモリ(703)上に記憶されているか否かを判断する。このために、LRU制御手段は、管理テーブル内に記録されているリンクテーブルを検索する(S203)。管理テーブルの一例は図5及び図6に示す。

【0036】管理テーブルは、ポインタテーブルとリンクテーブルとから成る。ポインタテーブルには図5

(A)に示すように、LRUの最上位と最下位に位置づけられているデータの論理アドレスが表わされている。リンクテーブルには図5(B)に示すように、キャッシュメモリ上に登録されているブロック毎に論理ブロックのリンク状態が示されている。このリンク状態は、直前の論理アドレスと直後の論理アドレスとを示すことで表わされている。エントリテーブルは、図6に示すように、対象ブロック番号と登録されているブロックを示すビットマップと、パリティデータ格納アドレスとから構成されている。

【0037】エントリテーブルの同一行には、同一のパリティグループを構成するブロックのキャッシュメモリ

への登録状況が記録されている。ここで、同一パリティグループを構成するブロックであっても、以前に読み出し要求を受けたか否かにより、キャッシュメモリ上に登録されているか否かが異なる。したがって、同一パリティグループを構成するブロックの内、どのブロックが登録されていて、どのブロックが登録されていないかを示すためにビットマップが使用される。

【0038】ビットマップは、4ビットから構成されており、各ビットが同一パリティグループを構成するブロックに対応している。ビットがセットされている(1になっている。)ブロックはキャッシュメモリ上に登録されている。また、ビットがセットされていない(0になっている。)ブロックはキャッシュメモリ上に登録されていないことを示している。

【0039】LRU制御手段(714)は、リンクテーブル内に対象データのエントリを発見すると、動作決定手段(713)に対し、対応するキャッシュメモリ上の格納アドレスとともにその旨を通して処理を動作決定手段に戻す。動作決定手段は、制御部に対して、当該データがキャッシュメモリに登録されている旨及び当該データのキャッシュメモリ上の格納アドレスを通知する(S254)。このように、対象データがキャッシュメモリ上に登録されている状態を「対象データはキャッシュヒットした」という。対象データがキャッシュヒットした場合は、キャッシュ管理部の処理は以上で終了する(S255)。動作決定手段は再び制御部からの次の通信待ち状態に移る(S201)。LRU制御手段が、リンクテーブル内に対象データのエントリを発見できなかった場合は、動作決定手段にその旨を通して処理を動作決定手段に戻す。動作決定手段は、制御部に対して、当該データがキャッシュメモリに登録されていない旨の通知を行う(S204)。このように、対象データがキャッシュメモリ上に登録されていない状態を「対象データはキャッシュミスした」という。対象データがキャッシュミスした場合は、動作決定手段は、対象データをディスク装置からキャッシュメモリに読み出す必要があると判断する(S205)。そして、データ登録手段(715)を起動する。

【0040】データ登録手段は、制御部(701)から通知された論理アドレスを物理アドレスに変換すること、で、対象データが格納されているディスク装置及び当該ディスク装置内のブロック番号を求める。データ登録手段は、対象となるデータをキャッシュメモリ上に読み込む処理に移る。この動作を図3のフローチャートを用いて詳細に説明する。

【0041】データ登録手段(716)は、アクセス対象となるブロックを前記の計算により求める。この後、データ登録手段は、当該ブロックに格納されているデータを格納するための領域をキャッシュメモリ上に割当てて、キャッシュメモリ上への領域の割当ては、データ登

録手段が管理テーブルにエントリを確保することから始められる。

【0042】データ登録手段(715)は、まず、キャッシュメモリ(703)上に新たなデータを格納するための領域があるか否か及び各テーブル上に新たなエントリ用を登録するための領域が存在するか否かを確認する(S301)。キャッシュメモリ等に空き容量が存在しない場合は新たなデータの登録ができない。この場合は、データ登録手段は、キャッシュメモリ等を解放するために制御をデータ追出し手段に移す(S352)。

【0043】データ追出し手段はデータを登録するのに十分な領域をキャッシュメモリから解放する。これを「追出し処理」という。追い出し処理については後述する。データ追出し手段は、キャッシュメモリを解放した後、制御をデータ登録手段に戻す。データ登録手段はキャッシュメモリへのデータの登録処理を続行する。ここで、対象となるデータが属するパリティグループを構成する他のデータが、一つもキャッシュメモリ上に登録されていない状態であれば、データ登録手段は当該ブロックのデータだけでなく当該ブロックが属するパリティグループのパリティデータをもキャッシュメモリ上に登録する必要がある。このことを確認するために、データ登録手段は、グループ登録判定手段(717)を起動する。

【0044】グループ登録判定手段は、対象データが属するパリティグループを構成する他のいずれかのデータが既にキャッシュメモリ上に登録されているか否かを判断する。このために、グループ登録判定手段は管理テーブル内のエントリテーブルを参照する。グループ登録判定手段は、エントリテーブルの「対象ブロック」の中からアクセス対象となっているデータの論理ブロック番号を検索する(S302)。

【0045】グループ登録判定手段は検出結果をデータ登録手段に通知した後に制御をデータ登録手段に戻す。対象となる論理ブロックが検出された場合は、対象データが属するパリティグループのパリティデータは既にキャッシュメモリ上に登録されていることを意味する。また、対象となる論理ブロックが検出されなかった場合はキャッシュメモリ上に登録されていないことを意味する。パリティデータが登録されていない場合、データ登録手段は対象データの登録と同時にパリティデータの登録も行う必要がある。

【0046】データ登録手段は、対象ブロックがエントリテーブルの中から検出されなかった場合には、キャッシュメモリ上に格納しようとしているデータを登録するための1ブロック分の領域をキャッシュメモリ上に確保する。そしてデータ登録手段は、更に1ブロック分の領域をキャッシュメモリ上に確保する必要がある(S304)。これは、今回のアクセス対象となっているブロックが含まれるパリティグループに属するパリティデータ

をもキャッシュメモリ上に登録するためである。データ登録手段は、キャッシュメモリへデータを登録するために確保した2ブロック分の領域を、アクセス対象となっているデータを含むブロックとパリティデータを含むブロックとに割当てエントリテーブルに記録する(S305)。

【0047】データ登録手段は、今回登録されるブロックのパリティグループ内の位置を計算により求める。そして、データ登録手段は、ビットマップを作成して(S306)エントリテーブルに記録し(S307)、エントリテーブルに対する新規エントリの登録を完成する。一方、対象ブロックがエントリテーブルの中から検出された場合には、データ登録手段は、キャッシュメモリ上に格納しようとしているデータを登録するための1ブロック分の領域のみをキャッシュメモリ上に確保する(S354)。

【0048】データ登録手段は、キャッシュメモリへデータを登録するために確保した1ブロック分の領域を、アクセス対象となっているデータを含むブロックに割当て、エントリテーブルに記録する(S355)。データ登録手段は、今回登録されるブロックのパリティグループ内の位置を計算により求める。そして、データ登録手段は、既に登録されているビットマップの対応ビットをセットする(S356)。以上によりエントリテーブルに登録されている既存エントリの更新が完了する。

【0049】次に、データ登録手段は、管理テーブルの更新を行う。データ登録手段は、新規に作成したエントリをLRUリンクの最上位に登録する(S308)必要が有る。このため、図5に示したリンクテーブルとポインタテーブルを更新する。まず、データ登録手段は、ポインタテーブルを参照し、現在の最上位論理ブロックを把握する。次に、リンクテーブルの新規エントリに対応する登録を有効にする。このため、データ登録手段は、ポインタテーブルにForwardポインタには無効値が代入され、Backwardポインタには現在の最上位ポインタが代入されたエントリを新規に登録する。その後、データ登録手段は、ポインタテーブル上の今回最上位の地位を譲った論理ブロックのエントリに記録されているForwardポインタを新たに今回最上位に位置づけられた論理アドレスに変更する(S309)。

最後に、データ登録手段は、対象データ及び必要によってはパリティデータを含むディスク装置に対してキャッシュメモリ上へのデータ転送を指示する(S310)。

【0050】以上により、上位装置から指定されたデータをキャッシュメモリ上に格納する。また、対象データのキャッシュメモリへの登録がパリティグループ内の最初の登録である場合には、対象データとともに当該データを格納しているブロックが属するパリティグループのパリティデータもキャッシュメモリ上に格納することとなる。

【0051】なお、本実施例においては、リンク管理テーブルの更新をキャッシュメモリへのデータの登録前に行うこととしている。しかし、これはキャッシュメモリへのデータの登録を確認した後にリンクテーブルを更新するという手段を排除するものではない。データ登録手段は、ディスク装置に対してデータの転送を指示すると、制御を動作決定手段に戻す（S311）。以上で図3の説明は終了する。以下、説明を図2のフローチャートに戻して継続する。

【0052】動作決定手段（713）は、ディスク装置からキャッシュメモリ上へのデータの転送終了を待つ（S206）。キャッシュ管理部は、データ転送が終了した後、制御部に対しデータの登録が完了した旨及び当該データのキャッシュメモリ上の格納アドレスを通知する（S207）。以上でキャッシュ管理部の処理は終了する（S208）。キャッシュ管理部は再び制御部からの次の通信待ち状態に移る（S201）。

【0053】さて、ここからは前述したように追出し処理について説明する。追出し処理とは、データ登録手段が、キャッシュメモリ上に新規のデータを格納するための領域を確保できない場合に、既にキャッシュメモリ上に登録されている一部のデータを未登録状態に変更することをいう。データ登録手段は、データ追出し手段（716）を起動して追出し処理を行う。

【0054】以下、図4のフローチャートを用いて、この追出しの動作を詳細に説明する。データ追出し手段は、管理テーブル内に存在する図3（A）に示したポインタテーブルを参照する。データ追出し手段は、LRUリンクの最下位に位置づけられている論理アドレスを認識する（S401）。キャッシュメモリからの追出しの対象となるのは、この論理ブロックである。データ追出し手段は、リンクテーブルの中から追出し対象として決定された論理ブロックを検索する（S402）。データ追出し手段は、該当エントリのForwardポインタの値を認識する（S404）。この値に対応する論理ブロックが、この後に最下位に位置づけられる論理ブロックとなる。

【0055】データ追出し手段は再びリンクテーブルを検索し、この後に最下位に位置づけられる論理ブロックのエントリを検出する（S405）。データ追出し手段は、該当エントリを検出した後backwardポインタに無効値を代入する（S407）。データ追出し手段は、追出し対象論理ブロックの登録を無効とする。このために、データ追出し手段は、追出し対象ブロックが登録されていたエントリに無効値を代入する（S408）。次に、データ追出し手段は、エントリテーブルから追出し対象のブロックに関する情報を削除する処理を開始する。

【0056】データ追出し手段は、追出し対象となっている論理ブロック番号をエントリテーブルの「対象ブ

ロック番号」の項から検索する（S409）。データ追出し手段は、目的のエントリを検出した後、当該エントリのビットマップを更新する。具体的には追出し対象となっている論理ブロックのパリティグループ内の位置を求め、対応するビットをリセットする（S411）。次に、データ追出し手段はグループ登録判定手段（717）を起動する。グループ登録判定手段は、追出し対象となった論理ブロックの属するパリティグループを構成する他のデータがキャッシュメモリ上に登録されているか否かを判定する。

【0057】グループ登録判定手段は、データ追出し手段がリセットした後のビットマップを参照する。そして、これらの全てが0となっているか否かを判定する（S412）。ビットマップが全て0であれば、当該パリティグループに属するデータはキャッシュメモリ上には一切存在しないことを意味する。逆に、一つでも0以外のビットが存在するのであれば、当該パリティグループに属するデータがキャッシュメモリ上に未だ存在していることとなる。

【0058】グループ登録判定手段は、追出し対象となったデータが属するパリティグループを構成する他のデータが、キャッシュメモリ上に存在するか否かという結果を伴って、制御をデータ追出し手段に戻す。追出し対象となったデータが属するパリティグループを構成する他のデータが、キャッシュメモリ上に存在しない場合は、データ追出し手段はキャッシュメモリを効率的に利用するためにパリティデータも追い出すこととする（S413）。パリティデータの追出し処理は、エントリテーブルに関する処理を除いて前述した通常データの追出し処理と同様に行われる。エントリテーブルに関しては、データ追出し手段が、該当エントリの「対象ブロック」の項に無効値を記録して、エントリ自体を解放することとなる（S414）。

【0059】一方、追出し対象となったデータが属するパリティグループを構成する他のデータがキャッシュメモリ上に存在する場合であれば、データ追出し手段は、前述したビットマップの更新のみを行い、パリティデータはキャッシュメモリ上に維持し、エントリテーブル内のエントリも削除しない。これで、データ追出し処理は終了し（S415）、データ追出し手段は制御をデータ登録手段に戻すこととなる。これにより、キャッシュメモリ上から、少なくとも1ブロック分の領域が解放されることとなる。

【0060】さて、以上の説明においては1論理ブロックを単位に追出し制御を行っている。しかし、制御装置の性能及びキャッシュメモリの容量いかによっては、追出し処理を1パリティグループ単位に行った方が妥当な場合もある。この場合には、例えばリンクテーブルのエントリをパリティグループ毎に設計し、ポインタテーブルが示す値をエントリ番号に変更することで対

応できる。

【0061】なお、本実施の形態においては、キャッシュ管理部(710)はMPU(711)とメモリ(712)とを備えることとし、各機能はMPUがメモリに格納されたマイクロプログラムに従って動作することによって実現されるものとして説明した。しかし、これはキャッシュ管理部に備える各機能をハードワイヤードロジックで実現し、本発明と同様の機能を実現するものを排除するものではない。

C. 制御部の発明

図10は、制御部(701)の動作を示したフローチャートである。制御部は、通常、上位装置(730)からのアクセス又はキャッシュ管理部(710)からの応答待ちの状態にある(S1001)。制御部は、いずれかのアクセスがあったことを検出すると(S1002)、当該アクセスが上位装置からのものなのか、あるいはキャッシュ管理部からのものなのかを判断する(S1003)。

【0062】アクセスが上位装置からのものであった場合は、制御部はキャッシュ管理部に対象となるデータがキャッシュメモリ上に登録されているか否かを問い合わせる(S1005)。制御部はキャッシュ管理部からの応答を待つ(S1006)。制御部は、キャッシュ管理部からの応答を解析し、対象データがキャッシュヒットなのかキャッシュミスであったのかを判断する(S1007)。キャッシュヒットであった場合は、キャッシュ管理部からキャッシュメモリ上における対象データの格納アドレスも通知される。制御部は、当該アドレスと上位装置との間でデータ転送を開始する(S1051)。これにより、上位装置から要求された一連の処理は終了する(S1052)。制御部は再びアクセス待ち状態(S1001)に復帰する。

【0063】キャッシュミスであった場合は、制御部は対象データの読み出し処理が完了するのを待つ。このため、制御部は処理を一旦中断する(S1008)。制御部は、後にキャッシュ管理部からの応答を検出した場合に、中断された処理を再開することとなる(S1003)。制御部は、この間全く別の処理を行うことが可能である。

【0064】制御部がキャッシュ管理部からの応答を検出すると(S1003)、当該応答とともに通知されるキャッシュメモリと上位装置との間でデータ転送を行う(S1051)。以上で制御部は処理を終了する(S1052)。これにより、上位装置から要求された一連の処理は終了することとなるので、制御部は再びアクセス待ち状態(S1001)に復帰する。

【0065】

【実施例】次に、入出力制御装置が行う各処理ごとに、制御部及びキャッシュ管理部が具体的にどのように連携して動作するかを説明する。

A. キャッシュに登録するデータをブロック単位に管理する場合の実施例

(1) キャッシュヒットの場合

図11に示すオペレーションフロは、読み出し要求対象のデータがキャッシュメモリ上に記憶されている場合の制御部(701)及びキャッシュ管理部(710)の動作を示している。制御部は、上位装置(730)からの読み出し要求を受け付けると、対象データがキャッシュメモリ(703)上に存在するか否かを判断するためにキャッシュ管理部にその旨を問い合わせる(S1101)。この問い合わせは、上位装置から指定される論理アドレスを伴って行われる。キャッシュ管理部は、制御部から通知されたデータがキャッシュメモリ上に記憶されているか否かを判断する。このために、キャッシュ管理部は管理テーブル内のリンクテーブルを参照する。管理テーブルに備えられる各テーブルは図5及び図6に示す。管理テーブルの構成に関しては既に説明したので、ここでは省略する。

【0066】図5(A)に示すポインタテーブルにおいては、リンクの最上位に位置づけられている論理ブロックがLB-13であり、リンクの最下位に位置づけられているブロックがLB-3であることを示している。図5(B)に示すリンクテーブルにおいては、論理ブロック13は最上位であるから直前の論理ブロックを表わすポインタに無効値が登録されている。また、エントリ3は最下位であるため直後のエントリを表わすポインタに無効値が登録されている。

【0067】例えば、LB-3、LB-18、LB-4～LB-7、LB-13の順でデータがキャッシュに登録された場合は、ポインタテーブル及びリンクテーブルはい上りになる。またエントリテーブルは図6に示すようになる。エントリテーブル内のビットマップは、登録されている論理ブロックに対応するビットのみセットされる。したがって、それぞれ、0001、0010、11100100、0000となる。なお、本実施例にあつては、1パリティグループに含まれる、上位装置との間で転送されるデータを格納しているブロックは4ブロックであり、かつ、論理アドレスの配置は昇順である。したがって、図6に示すエントリテーブルの「対象ブロック番号」には、パリティグループの先頭論理アドレスのみ記録しておくことで十分である。例えば上位装置から指定された論理アドレスが11であれば、「対象ブロック番号」に「8」と示されているエントリを探し出せばよい。

【0068】さて、本実施例にあつては、キャッシュメモリ上に対象データが登録されていると仮定している。したがって、キャッシュ管理部は、エントリテーブルを参照した結果、キャッシュメモリ上に当該データが存在すると認識する(S5201)。例えば、上位装置から指定された論理アドレスが13であるとする。キャッシュ

キャッシュ管理部は、エントリテーブルを参照してエントリ3（上から三つ目のエントリ）に対象ブロックが属しているパリティグループが登録されていることを発見する。次に、キャッシュ管理部はビットマップを参照して当該論理ブロックが登録されていることを認識する。

【0069】キャッシュ管理部は、対象データがキャッシュメモリ上に記憶されている旨を当該データが格納されているメモリ上のアドレスと共に制御部に通知する（S1152）。これでキャッシュ管理部の処理は終了する（S1153）。制御部は、キャッシュ管理部から通知されたメモリアドレスを参照して、上位装置から要求されてデータを送出する（S1102）。これで制御部の処理は終了する（S1103）。

(2) キャッシュミスの場合

図12のオペレーションフローは、読み出し要求対象のデータが、キャッシュメモリ（703）上に記憶されていない場合の制御部（701）及びキャッシュ管理部（710）の動作を示している。

【0070】制御部は、上位装置からの読み出し要求を受け付けると、対象データがキャッシュメモリ上に存在するか否かを判断する。このために、制御部はキャッシュ管理部にその旨を問い合わせる（S1201）。この問い合わせは、キャッシュヒットの実施例で示した場合と同様に論理アドレスを伴って行われる。キャッシュ管理部は、制御部から通知されたデータがキャッシュメモリ上に記憶されているか否かを判断する必要がある。このために、キャッシュ管理部はキャッシュヒットの実施例で説明したエントリテーブルを参照する。

【0071】本実施例においては、キャッシュメモリ上に対象データが存在しないと仮定している。したがって、キャッシュ管理部はエントリテーブルを参照した結果、対象となる論理アドレスを発見できない（S1251）。具体的には、キャッシュ管理部はリンクテーブル図5（B）を参照して、テーブル内に当該論理ブロックが登録されているか否かを判断する。例えば、上位装置が指定した論理アドレスが11であって、リンクテーブルが図5（B）に示す状態であったとすると、論理ブロック11は、登録されていないこととなる。

【0072】キャッシュ管理部は、対象データがキャッシュメモリ上に記憶されていない旨及び当該データをディスク装置からキャッシュメモリ上に読み込む処理に移行する旨を制御部に通知する（S1252）。対象データがキャッシュメモリ上に記憶されていない旨の通知を受けた制御部は、キャッシュ管理部が対象データをキャッシュメモリ上に記憶するまで上位装置から受け付けた処理を一時中断する（S1202）。そして、制御部はデータの登録が完了するまで別の処理を進める。

【0073】一方、キャッシュ管理部は、当該データをキャッシュメモリ上に読み込むための動作に移る。キャッシュ管理部は、制御部から通知された論理アドレス

を物理アドレスに変換する。これにより、キャッシュ管理部は対象データが格納されているディスク装置及び当該ディスク装置内のブロック番号を求める。キャッシュ管理部は、当該論理ブロックのデータをキャッシュメモリ上に読込む。また、キャッシュ管理部は、必要によっては当該ブロックが属するパリティグループに含まれるパリティデータデータをもキャッシュメモリ上に読み込む（S1253）。この具体的な動作を、図3のフローチャートと対比させながら説明する。

【0074】キャッシュ管理部は、論理アドレスを前述した計算式及び図9を利用して物理アドレスに変換する。例えば、上位装置が指定した論理アドレスが11であったとすると、値A=3、値B=2、値C=2となるから物理アドレスはDISK4/BLOCK2となる。キャッシュ管理部は、アクセス対象となるブロックを前記の計算により求めた後、当該ブロックに格納されているデータを格納するための領域をキャッシュメモリ上に割当てて、

【0075】キャッシュメモリ上への領域の割当ては、管理テーブルにエントリを確保することから始められる。キャッシュ管理部は、リンクテーブルを参照し、エントリに空きがあるか否かを判断する（S301）。ここで、空きが無いと判断した場合はリンクの最下位のエントリを解放する（S352）。詳細は「データの追い出し」として、後述する。

【0076】この時のテーブルの状況が例えば図5の状態であったとすると、未使用領域が存在することとなる。したがって、キャッシュ管理部は新たな領域に論理ブロック11を登録することとする。エントリテーブルに新規エントリの登録が可能である場合、あるいは最下位のエントリを解放した場合は、キャッシュ管理部はキャッシュメモリ上に格納しようとしているデータを登録するための1ブロック分の領域をキャッシュメモリ上に確保する。キャッシュ管理部は、更に1ブロック分の領域をキャッシュメモリ上に確保する（S304）。これは、今回のアクセス対象となっているブロックが属するパリティグループのパリティデータをもキャッシュメモリ上に登録するためである。

【0077】例えば、上位装置が実際に必要としているデータはLB-11のみであったとしても、本実施例にかかるディスク制御装置では、当該データ（本実施例にあつてはLB-11）の属するパリティグループのパリティデータ（P2）をもキャッシュメモリ上に登録することになる。キャッシュ管理部は、キャッシュメモリへデータを登録するために確保した2ブロック分の領域を、アクセス対象となっているデータを含むブロックとパリティデータを含むブロックとに割当てて、キャッシュ管理部は、この情報をエントリテーブルに記録する（S305）。

【0078】キャッシュ管理部は、今回登録されるプロ

ックはパリティグループ内の三つ目のブロックに相当することを計算により求めることができる。したがって、キャッシュ管理部はビットマップとして0001を作成する。そして、キャッシュ管理部はエントリテーブルにこの情報を記録する。これで、キャッシュ管理部によるエントリテーブルに対する新規エントリの登録は完了する(S307)。

【0079】次に、キャッシュ管理部は、リンクテーブルの更新を行う。キャッシュ管理部は、新規に作成したエントリをLRUリンクの最上位に登録する(S309)必要が有る。このため、キャッシュ管理部はリンクテーブルとポインタテーブルを更新する。まず、キャッシュ管理部は、ポインタテーブルを参照して現在の最上位論理ブロックを把握する。次に、キャッシュ管理部は、リンクテーブルの新規エントリに対応する登録を有効にする。このため、キャッシュ管理部はForwardポインタには無効値を代入し、Backwardポインタには現在の最上位ポインタを登録する。その後、キャッシュ管理部は、今回最上位の地位を譲った論理ブロックのForwardポインタを新たに今回最上位に位置づけられた論理アドレスに変更する。

【0080】今回、新たに登録された論理アドレスが11であるとすれば、キャッシュ管理部はリンクテーブル上の論理ブロック11のForwardポインタを無効値(xx)に、Backwardポインタを13に変更する。更にキャッシュ管理部は、論理ブロック13のForwardポインタも11に変更する。また、キャッシュ管理部は、ポインタテーブルのトップポインタの値も13から11に変更する。次に、キャッシュ管理部は論理ブロック11を格納するキャッシュメモリ上のアドレスを記録する。これでキャッシュ管理部によるテーブルの更新が完成する。

【0081】最後に、キャッシュ管理部は、LB-11に対応するディスク装置(物理アドレスDisk4/Block2)及びP2を格納しているディスク装置(物理アドレスDisk2/Block2)に対し、リンクテーブル及びエントリテーブルに登録されたキャッシュメモリアドレスに対して、データを転送するように指示する。

【0082】このようにして、キャッシュ管理部は、アクセス対象となっているディスク装置のみならず、パリティデータを格納しているディスク装置に対しても、格納されているデータをキャッシュメモリ上に転送するように指示するのである。以上により、本発明における入出力制御装置は、上位装置から指定されたデータとともに当該データを格納しているブロックが属するパリティグループに含まれるパリティデータをキャッシュメモリ上に格納することができる。

【0083】なお、本実施例においては、リンク管理テーブルの更新はキャッシュメモリへのデータの登録前に行

われることとしている。しかし、これはキャッシュメモリへのデータの登録を確認した後にリンクテーブルを更新するという手段を排除するものではない。アクセス対象のデータ及びパリティデータのキャッシュメモリへの転送が終了すると、キャッシュ管理部は、上位装置から転送を要求されたブロックのデータが格納されているキャッシュメモリ上の領域とともに、データがキャッシュメモリ上に登録されたことを、制御部に通知する(S1254)。

【0084】制御部は、前記通知を受け取ると、実行中の処理を中断又は終了する。その後、制御部はキャッシュ管理部から通知されたメモリアドレスを認識し(S1203)、当該アドレスからデータを読み出す(S1204)。なお、アクセス対象となった論理ブロックのデータが、キャッシュメモリ上に記憶されていない場合であっても、エントリテーブルに当該ブロックの属すべきパリティグループのエントリが登録されている場合もある。これは、当該論理ブロックへのアクセスが行われる以前に同一パリティグループ内の他の論理ブロックへのアクセスがあった場合に生じる。

【0085】このような場合には、キャッシュ管理部は、リンクテーブルの更新を行った後、エントリテーブル上の対応するエントリに当該アクセス対象となったデータの情報を書き込むだけで良い。具体的には、キャッシュ管理部は、パリティグループ内のブロック位置を計算し、該当エリアにキャッシュメモリ上のアドレスを記録し、ビットマップを更新することとなる。

(3) キャッシュメモリからのデータの追い出し
キャッシュメモリ上に登録されているデータを追い出す場合の動作は、キャッシュ管理部内の動作に閉塞される。したがって、既に説明した内容と異なるところはないので、以下具体的な数値を例示しつつ、簡単に説明する。

【0086】キャッシュ管理部は、ポインタテーブルを参照してキャッシュメモリからの追い出しの対象となる論理ブロックを認識する。キャッシュ管理部は、リンクテーブルから追い出し対象のブロックを検索する。キャッシュ管理部は、追い出しブロックを発見したら、当該ブロックの登録を無効とする。また、キャッシュ管理部は当該ブロックのForwardポインタを参照して、次に最下位に位置づけられる論理ブロックを認識する。キャッシュ管理部は、再びリンクテーブルを検索して、次に最下位に位置づけられる論理ブロックのエントリを探し出す。キャッシュ管理部は、当該エントリのbackwardポインタに無効値を代入する。

【0087】次に、キャッシュ管理部は、エントリテーブルから追い出し対象のブロックに関する情報を削除する処理を開始する。キャッシュ管理部は、追い出し対象となっているブロック番号をエントリテーブルの「対象ブロック番号」の項から探し出す。キャッシュ管理部

は、目的のエントリを検出した後、リンクテーブルから対象ブロックを削除する。その後、キャッシュ管理部は、エントリテーブルの「対象ブロック」を参照して、削除対象となるブロックが含まれているエントリを探し出す。

【0088】キャッシュ管理部は、目的のエントリを探し出した後、ビットマップを更新する。本実施例にあつては、追い出し対象の論理アドレスを3としている。したがって、キャッシュ管理部は現在のビットマップと1110との論理積を新たなビットマップとする。

【0089】論理積を求めた結果、ビットマップが0000となれば、当該パリティグループに属するデータはキャッシュメモリ上に存在しないことを意味する。したがって、キャッシュ管理部はキャッシュメモリを効率的に利用するためにパリティデータも追い出す。そして、キャッシュ管理部は当該エントリの「対象ブロック」の項に無効値を記録し、エントリも解放する。

【0090】ビットマップが0000以外であれば、当該パリティグループに属するデータのいずれかがキャッシュメモリ上に存在していることとなる。したがって、キャッシュ管理部はビットマップの更新のみを行う。キャッシュ管理部は、パリティデータをキャッシュメモリ上に維持し、エントリも削除しない。

B. パリティグループを一単位として管理する場合の実施例

(1) キャッシュヒットの場合

図11に示すオペレーションフローは、読み出し要求対象のデータがキャッシュメモリ上に記憶されている場合の制御部及びキャッシュ管理部の動作を示している。

【0091】制御部は、上位装置からの読み出し要求を受け付けると、まず対象データがキャッシュメモリ上に存在するか否かを判断する必要がある。このために、制御部はキャッシュ管理部にその旨を問い合わせる(S1101)。この問い合わせは、上位装置から指定される論理アドレスを伴って行われる。これは前述した実施例Aの場合と同様である。

【0092】キャッシュ管理部は、制御部から通知されたデータがキャッシュメモリ上に記憶されているか否かを判断する必要がある。このために、キャッシュ管理部は管理テーブル内のエントリテーブルを参照する。エントリテーブルの一例は図14に示す。エントリテーブルの構造は、実施例Aの場合と異なる。

【0093】エントリテーブルはエントリ番号と、対象ブロック番号と、キャッシュ格納アドレスとから構成されている。キャッシュメモリ上に登録されたデータは、登録された順にエントリテーブルに追加される。そして、各データには昇順にエントリ番号が割当てられる。図は、LB-0〜LB-3、LB-4〜LB-7、LB-8〜LB-B、LB-C〜LB-Fの順で登録され、それぞれ0〜3のエントリ番号が与えられたことを示し

ている。

【0094】本実施例にあつては、キャッシュメモリ上に対象データが存在すると仮定している。したがって、キャッシュ管理部は、前記エントリテーブルを参照した結果、対象となるブロックを含む有効なエントリが存在することを発見する。そして、キャッシュ管理部はキャッシュメモリ上に当該データが存在すると認識することとなる(S1151)。

【0095】以下、実施例Aの場合と同様であるので、説明は省略する。

(2) キャッシュミスの場合

図12のオペレーションフローは、読み出し要求対象のデータがキャッシュメモリ上に記憶されていない場合の制御部及びキャッシュ管理部の動作を示している。

【0096】制御部は、上位装置からの読み出し要求を受け付けると、対象データがキャッシュメモリ上に存在するかいなかを判断する必要がある。制御部は、このためにキャッシュ管理部にその旨を問い合わせる(S1201)。この問い合わせは、キャッシュヒットの実施例で示した場合と同様に論理アドレスを伴って行われる。

【0097】キャッシュ管理部は、制御部から通知されたデータがキャッシュメモリ上に登録されているか否かを判断する必要がある。このために、キャッシュ管理部はキャッシュヒットの実施例で説明した場合と同様にエントリテーブルを参照する。本実施例においては、キャッシュメモリ上に対象データが存在しないと仮定している。したがって、キャッシュ管理部は、前記エントリテーブルを参照した結果、対象となるデータを含む有効なエントリを発見できない(S1251)。

【0098】具体的には、キャッシュ管理部は、エントリテーブルを参照して、「該当ブロック」の項に当該論理ブロックが登録されているか否かを判断する。例えば、上位装置が指定した論理アドレスが11であつて、エントリテーブルが図14に示す状態であつたとすると、論理ブロック11は登録されていないこととなる。キャッシュ管理部は、対象データがキャッシュメモリ上に登録されていない旨及び当該データをディスク装置からキャッシュメモリ上に読み込む処理に移行する旨を制御部に通知する(S1252)。

【0099】対象データがキャッシュメモリ上に登録されていない旨の通知を受けた制御部は、キャッシュ管理部が対象データをキャッシュメモリ上に登録するまで、上位装置から受け付けた処理を一時中断する(S1202)。制御部は、この間別の処理を進める。キャッシュ管理部は、制御部から通知された論理アドレスを物理アドレスに変換する。これによりキャッシュ管理部は、対象データが格納されているディスク装置及び当該ディスク装置内のブロック番号を求める。キャッシュ管理部は、当該ブロックが属するパリティグループの全てのデータをキャッシュメモリ上に読み込む(S1253)。この

動作を図15に示したフローチャートを用いて説明する。

【0100】キャッシュ管理部は、論理アドレスを前述した計算式及び図9を利用して物理アドレスに変換する。例えば、上位装置が指定した論理アドレスが11であったとすると、前述した通り、値A=3、値B=2、値C=2となる。したがって物理アドレスはDISK4/BLOCK2となる。キャッシュ制御装置は、アクセス対象となるブロックを前記の計算により求める。その後、キャッシュ管理部は当該ブロックに格納されているデータを格納するための領域をキャッシュに割り当てる。

【0101】キャッシュメモリ上への領域の割当ては、エントリテーブルにエントリを確保することから始められる。キャッシュ管理部は、エントリテーブルを参照し、新規の登録が可能であるか否かを判断する(S1501)。キャッシュ管理部は、空きが無いと判断した場合はリンクの最下位のエントリを追出す(S1551)。この処理は前述した追出し処理の項で説明したものである。。

【0102】この時のテーブルの状況が例えば図14の状態であったとすると、図ではエントリ4が未使用であるので、キャッシュ管理部は新たな領域をエントリ4として登録することとする。エントリテーブルに新規エントリの登録が可能である場合、あるいは最下位のエントリを解放した場合は、キャッシュ管理部は、今回対象となっているデータを登録するための1ブロック分の領域をキャッシュメモリ上に確保する。また、キャッシュ管理部は、更に4ブロック分の領域をキャッシュメモリ上に確保する(S1502)。これは、キャッシュ管理部が、今回のアクセス対象となっているブロックが含まれるパリティグループに属する全てのブロックのデータをキャッシュメモリ上に登録するためである。

【0103】例えば、上位装置が実際に必要としているデータはLB-11のみであったとすると。この場合でも、本発明にかかるディスク制御装置は、当該データ(本実施例にあってはLB-11)の属するパリティグループを形成している残りのデータの全て(LB-8、LB-9、P2、LB-10)をキャッシュメモリ上に登録することになる。

【0104】キャッシュ管理部は、キャッシュメモリへデータを登録するために確保した5ブロック分の領域を、パリティグループを構成するそれぞれの論理ブロックに割り当てる(S1503)。そしてキャッシュ管理部は、その情報をエントリテーブルに記録する。これにより、エントリテーブルに対する新規エントリの登録は完成する(S1504)。

【0105】キャッシュ管理部は、新規に作成したエントリをLRUリンクの最上位に登録する(S1506)。このため、キャッシュ管理部はリンクテーブルと

ポインタテーブルを更新する。ポインタテーブルを図13(A)にリンクテーブルを図13(B)に示す。まず、キャッシュ管理部は、ポインタテーブルを参照して現在の最上位エントリを把握する。次に、キャッシュ管理部は、リンクテーブルに登録する新規エントリを有効にする。このために、キャッシュ管理部はForwardポインタには無効値を代入し、Backwardポインタには現在の最上位エントリを登録する。その後、キャッシュ管理部は、今回最上位の地位を譲ったエントリのForwardポインタを新たに今回最上位に位置づけられたエントリ値に変更する。例えば今回新たに登録されたエントリが4であるとする。この場合、キャッシュ管理部は、リンクテーブル上のエントリ4のForwardポインタを無効値(xx)に、Backwardポインタを3に変更する。更に、キャッシュ管理部はエントリ3のForwardポインタを4に変更する。また、キャッシュ管理部はポインタテーブルのトップポインタの値を3から4に変更する。これで、テーブルの更新は完成する。

【0106】最後に、キャッシュ管理部は、各ディスク装置に対し、エントリテーブルに登録されたキャッシュメモリアドレスに対して、データを転送するように指示する。つまり、キャッシュ管理部は、パリティグループを構成する全てのディスク装置に対して、データをキャッシュメモリ上に格納するように指示するのである。

【0107】なお、本実施例においては、リンクテーブルの更新は、キャッシュメモリへのデータの登録前に行うこととしている。しかし、これはキャッシュメモリへのデータの登録を確認した後にリンクテーブルを更新するという手段を排除するものではないことは、実施例Aの場合と同様である。以上で、パリティグループを構成するブロックに含まれる全てのデータをキャッシュメモリ上に登録する処理が終了する。

【0108】以下、制御部のデータ転送についての動作は実施例Aと同様であるので省略する。

(3) キャッシュメモリからのデータの追出し
本実施例におけるデータの追出しは、前述した実施例Aのデータの追出しと同様に制御されるので、説明は省略する。

【0109】

【発明の効果】請求項1、請求項4、請求項5及び請求項8に記載された発明にあっては、データがキャッシュメモリ上に登録されている限りパリティデータをキャッシュメモリ上に維持することができる。このため、キャッシュメモリ上に登録されているデータに対応するパリティデータは、必ずキャッシュメモリ上に存在することとなる。したがって、キャッシュメモリ上に登録されているデータに対する書込み処理は、全ての動作をキャッシュメモリ上で処理することができるので、大幅な処理速度の向上が図れる。

【0110】請求項2に記載された発明にあっては、既にキャッシュメモリに登録されているパリティデータについては、再度のデータ転送を抑止することができる。したがって、パリティデータのキャッシュメモリ上への登録及び追出しを最も効率よく行うことができる。すなわちパリティデータに関するディスクアクセスを最小回数に抑止することができるので、装置全体としての処理能力向上にも貢献する。

【0111】請求項3に記載された発明にあっては、パリティデータを一般のユーザデータと区別して取り扱うことができる。したがって、装置の保守性を向上させることができる。請求項6に記載された発明にあっては、予めパリティデータをキャッシュメモリ上に登録しておくことができる。このため、上位装置から初めての書き込み処理を要求された際にディスク装置へのアクセスを改めて行わないで済む。したがって、装置全体としての処理性能を維持しつつ、書き込み動作を高速に処理することが可能となる。

【0112】請求項7に記載された発明にあっては、あるパリティグループに含まれる全てのデータを一群のデータとして管理できる。したがって、各手段の構成を簡素化することが可能であり、処理速度の向上にも貢献する。繰り返して述べると、本発明によれば、あるデータがキャッシュメモリ上に登録されている限り、必ず当該データが属するパリティグループのパリティデータがキャッシュメモリ上に存在することとなる。

【0113】このため、データの書き込み処理が発生しても、ディスク装置からのパリティデータの読み込みが全く不要である。したがって、本発明にかかる入出力制御装置は、書き込みを高速に処理することが可能である。つまり、本発明によれば、上位装置が必要とするデータ（以下、「ユーザデータ」という。）に対応したパリティデータのみをキャッシュメモリ上に登録しつづけることとなる。このため、パリティデータの無駄な追出しや無駄な保持がなくなり、性能向上に貢献する。更にキャッシュメモリの有効利用も可能となる。これは、パリティデータとユーザデータを区別せずに単にキャッシュ管理をする装置では達成不可能な効果である。

【0114】また、本発明にかかる入出力制御装置は、書き込み処理又は読み出し処理に関わらず、あるパリティグループに対する最初のアクセスで、ホストとの間で転送されるデータとともにパリティデータをキャッシュメモリに登録する。これにより、ディスクへのアクセスが最小限に抑止されるため性能向上に寄与する。つまり、パリティデータは書き込みの際には必ず読み出す必要があるが、本発明によれば、書き込み処理とは同期させずにパリティデータをキャッシュメモリ上に格納することができるので、見かけ上ディスク待ち時間が削減可能となる。

【図面の簡単な説明】

【図1】本発明に係るキャッシュ管理部の構成図

【図2】キャッシュ管理部の制御動作全体を示すフローチャート

【図3】キャッシュ管理部のデータ登録動作を示すフローチャート

【図4】キャッシュ管理部のデータ追出し動作を示すフローチャート

【図5】管理テーブルの一例を示した図

(A) ポインタテーブル

(B) リンクテーブル

【図6】エントリテーブルの一例を示した図

【図7】入出力制御装置の全体構成図

【図8】ディスク装置のデータ配置の模式図

【図9】論理ブロック変換表

【図10】制御部の制御動作全体を示すフローチャート

【図11】キャッシュヒットの場合の、制御部とキャッシュ管理部との連携動作を示すオペレーションフロー

【図12】キャッシュミスの場合の、制御部とキャッシュ管理部との連携動作を示すオペレーションフロー

【図13】管理テーブルの一例を示した図

(A) ポインタテーブル

(B) リンクテーブル

【図14】エントリテーブルの一例を示した図

【図15】キャッシュ管理部のデータ登録動作を示すフローチャート

【図16】従来のキャッシュ管理部の構成図

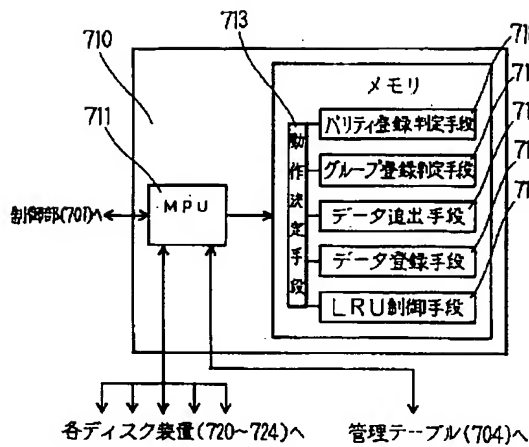
【図6】

対象ブロック	Bit Map	0Block	1Block	2Block	3Block	パリティ
0 LB-0~LB-3	0001				0x0003	0x0004
1 LB-16~LB-19	0010			0x0030		0x0005
2 LB-4~LB-7	1000	0x011				0x0015
3 LB-12~LB-15	0100		0x1029			0x102C
4						

【図9】

値C \ 値A	0	1	2	3	Parity
0	Disk0	Disk1	Disk2	Disk3	Disk4
1	Disk0	Disk1	Disk2	Disk4	Disk3
2	Disk0	Disk1	Disk3	Disk4	Disk2
3	Disk0	Disk2	Disk3	Disk4	Disk1
4	Disk1	Disk2	Disk3	Disk4	Disk0

【図1】



【図5】

Top	Bottom
13	3

(A)

エントリ	Forward	Backward
3	18	XX
18	4	3
4	13	18
13	XX	4

(B)

【図13】

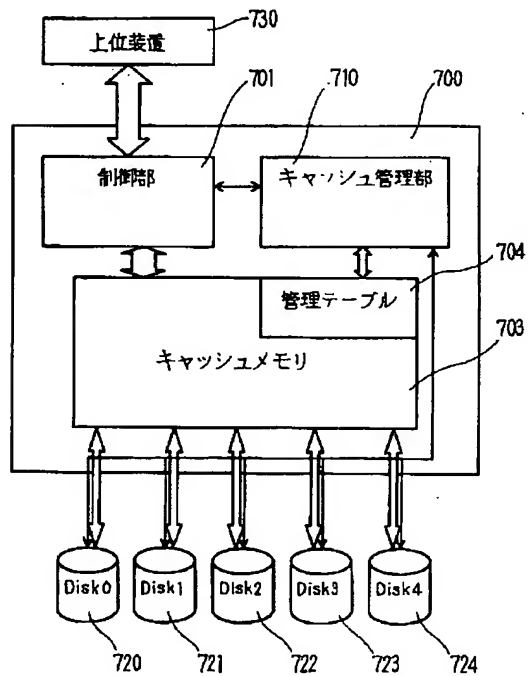
Top	Bottom
3	0

(A)

エントリ	Forward	Backward
0	1	XX
1	2	0
2	3	1
3	XX	2
4		

(B)

【図7】

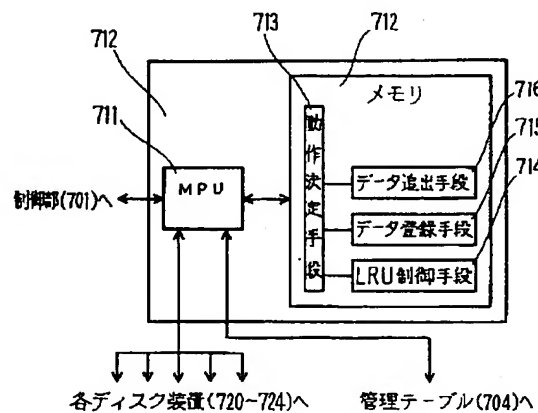


【図8】

LB-0	LB-1	LB-2	LB-3	P0	← Block0
LB-4	LB-5	LB-6	P1	LB-7	← Block1
LB-8	LB-9	P2	LB-10	LB-11	← Block2
LB-12	P3	LB-13	LB-14	LB-15	← Block3
P4	LB-16	LB-17	LB-18	LB-19	← Block4

Disk0 (720) Disk1 (721) Disk2 (722) Disk3 (723) Disk4 (724)

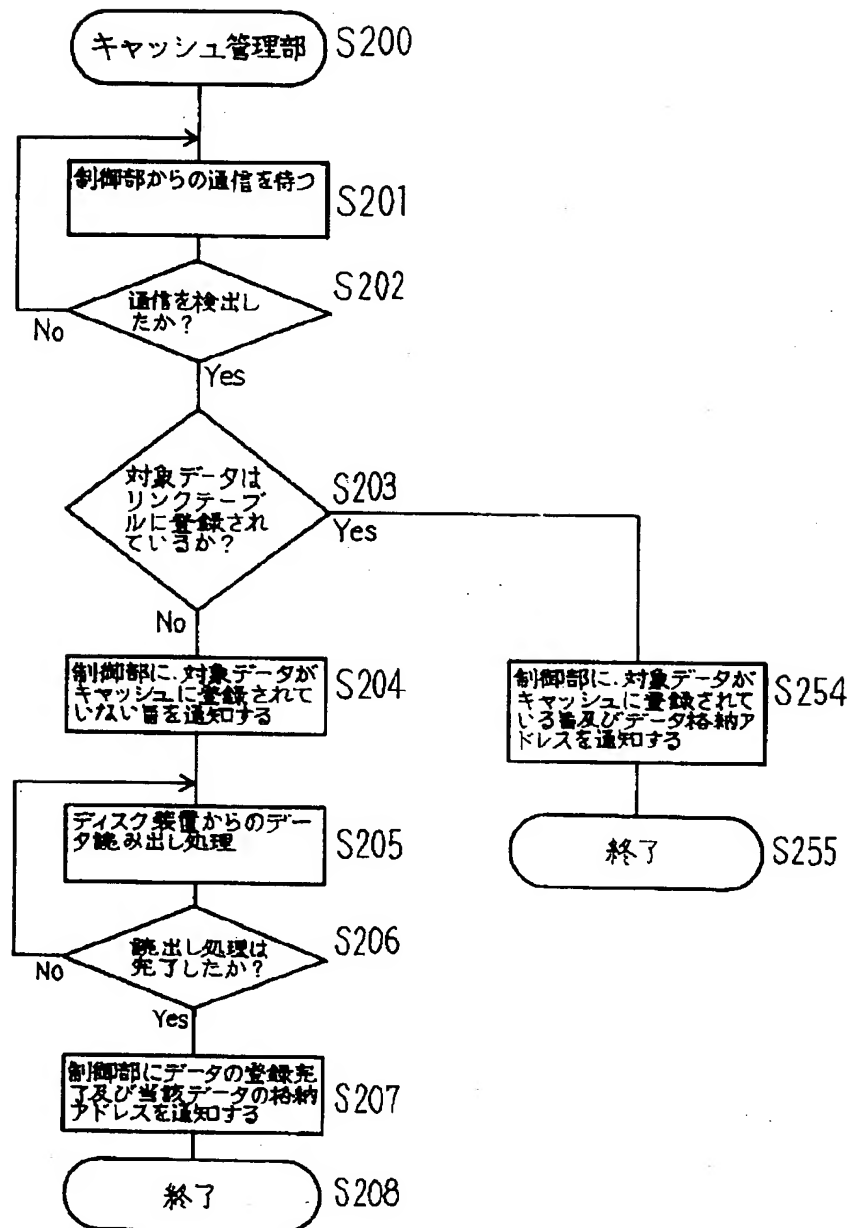
【図16】



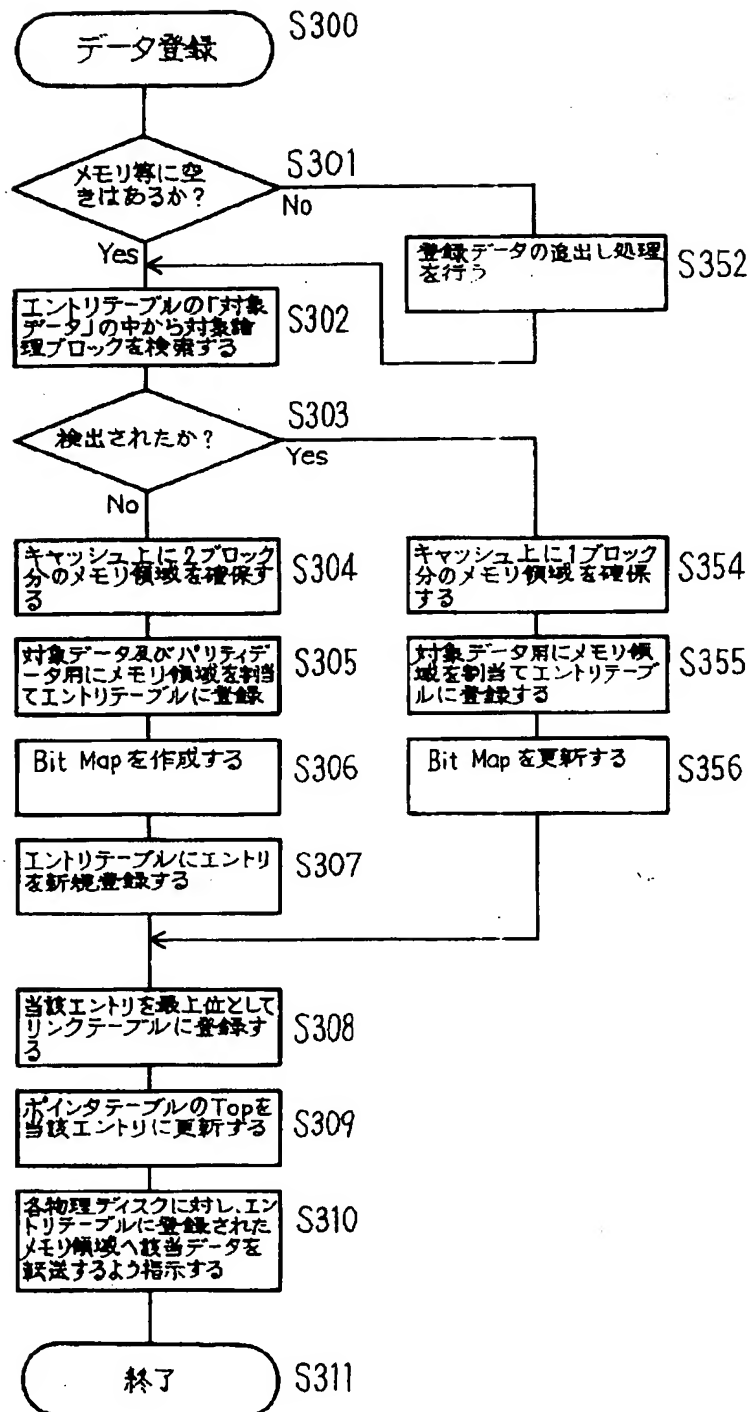
【図14】

	対象ブロック	0Block	1Block	2Block	3Block	パリティ
0	LB-0～LB-3	0x0000	0x0001	0x0002	0x0003	0x0004
1	LB-16～LB-19	0x0010	0x0021	0x0030	0x0034	0x0005
2	LB-4～LB-7	0x0011	0x0012	0x0013	0x0014	0x0015
3	LB-12～LB-15	0x1020	0x1029	0x102A	0x102B	0x102C
4						

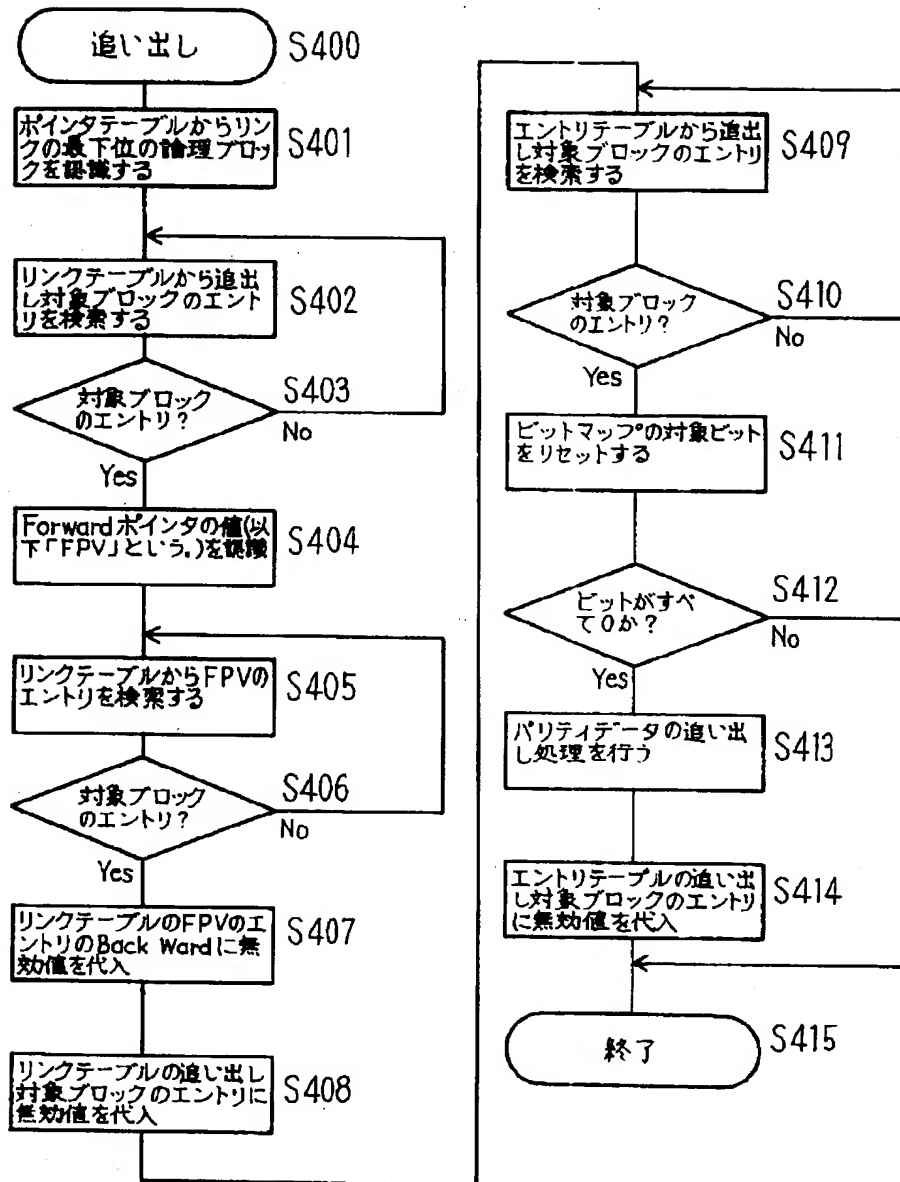
【図2】



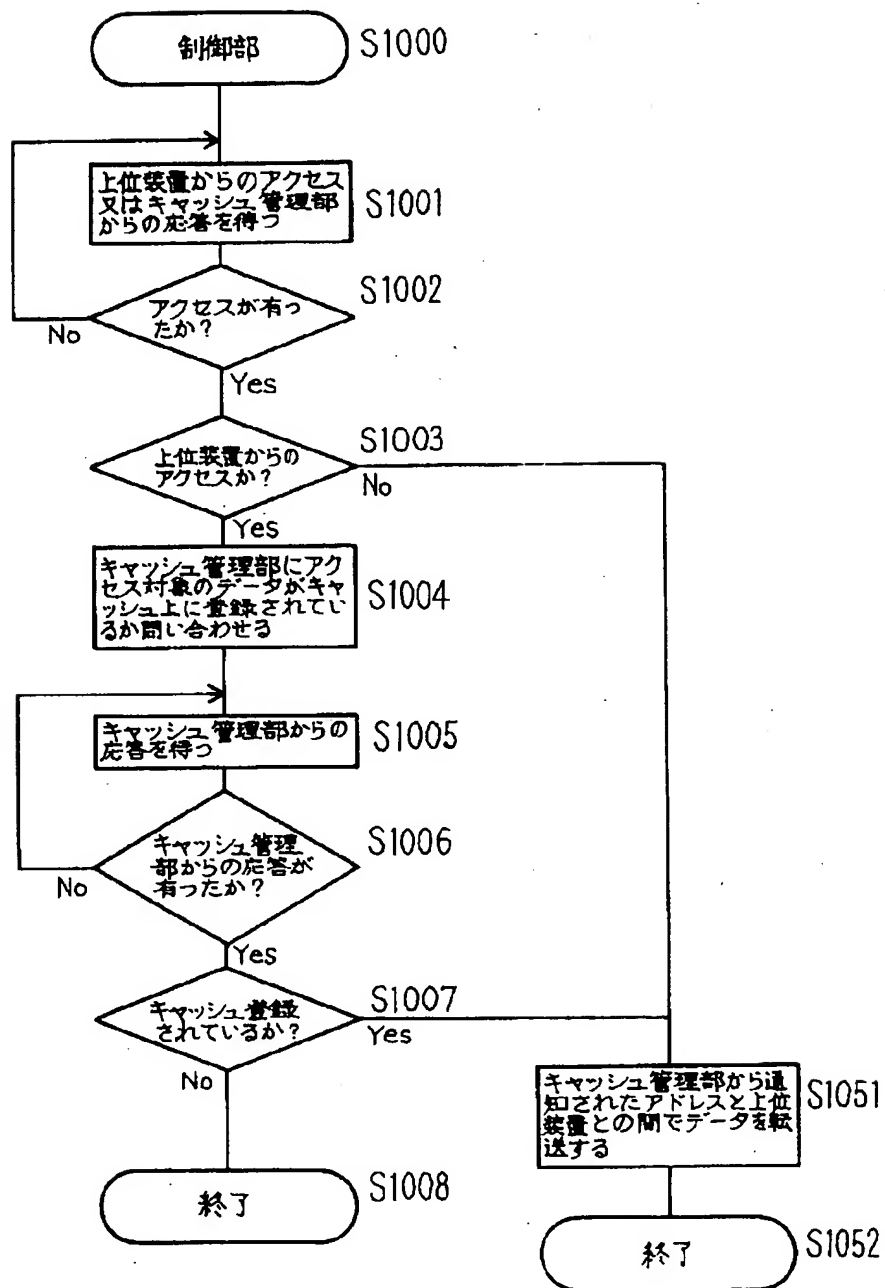
【図3】



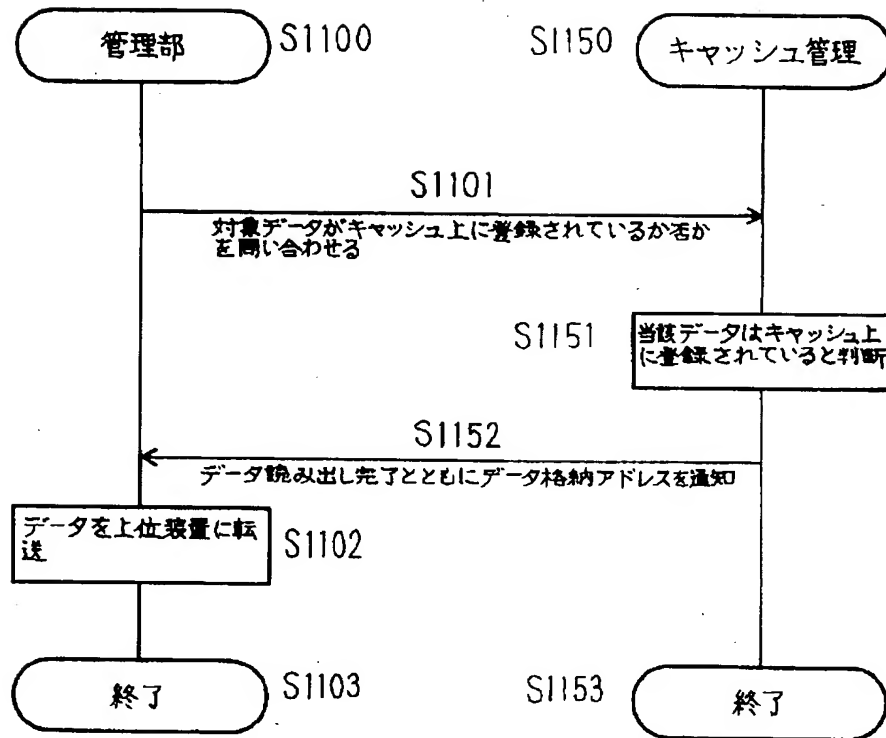
【図4】



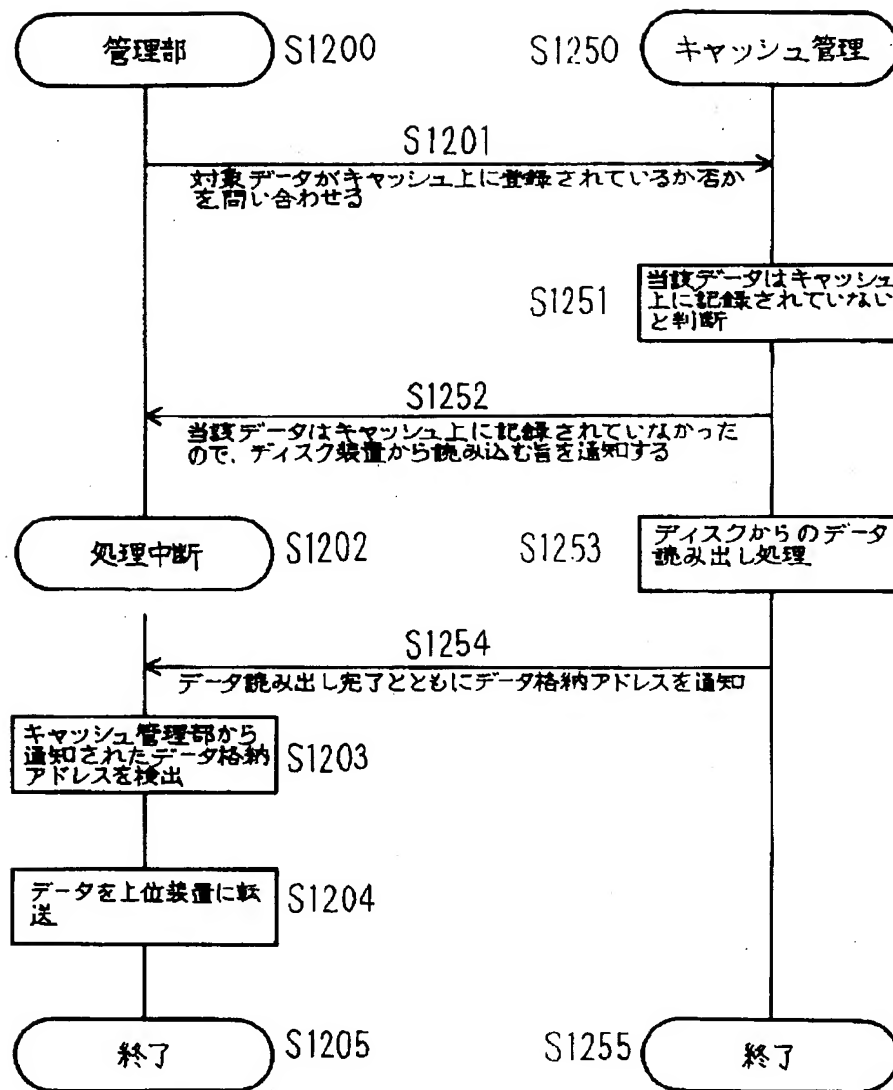
【図10】



【図11】



【図12】



【図15】

